Speech Processing 2015/2016

3rd Test

May 24th 2016

Please identify this form with your name and student number in the reserved spaces at the beginning and end of the test. Wrong answers to True-False questions will be penalized.

Name:	
Number:	
Group number:	

PART I: Speech Pattern Classification (Chapter 5)

1. (1.5) In ML, what is the difference between identification, verification and regression tasks? Provide an example from speech processing for each task.

	Example
Identification	
Verification	
Regression	

2. (1.6) There are many ways of dealing with the time-varying speech input nature in speech pattern classification tasks. Mention two approaches, indicating for each one if it is applied at the feature, model or output level.

3. (1.5) Consider the following feature vector sequence of static cepstral coefficients:

Time	f1	f2	f3
t=1	3	-2	7
t=2	2	-4	5
t=3	4	0	3

(a) First, apply cepstral mean normalization (CMN) to the feature vector sequence:

Time	f1	f2	f3
t=1			
t=2			
t=3			

(b) Then, compute the vector of static + first order delta features at time t=2 of the previous resulting features:

Time	f1	f2	f3	df1	df2	df3
t=2						

4. (1.5) Assume that in a 2-class classification problem in speech, each class is modelled with a Gaussian mixture model (GMM) of 64 mixtures with diagonal covariance. The dimensionality of the feature space is 13. What is the total number of parameters? Indicate the meaning of each quantity in your computation.

5. (1.0) What does GMM-UBM stand for?

____

6. (1.0) Complete the sentence:
____________________ are vectors of stacked Gaussian means.

PART II: Automatic Speech Recognition (Chapter 6)

7. (1.5) Consider the HMM models developed in the lab for isolated word speaker dependent recognition (names of months plus silence). How many transition probabilities need to be re-estimated at each iteration of the Baum -Welsh algorithm for the total number of models? Start by specifying an appropriate number of states, and a common model configuration for all word models. The selected configuration should minimize the number of transitions to be estimated.

8. (1.0) What is the part of the syllable which is less affected by a significant increase of speaking rate?

____

9. (1.5) Consider the language models for the following domains:

   (a) Time schedule of the matches of the final phase of EURO 2016

   (b) isolated digit recognition

   (c) general purpose newspaper text

   (d) weather forecast for Portugal

   Reorder them by increasing order of perplexity (e.g. $(a) < (b) < (c) < (d)$).

   ____

10. (2.5) Consider the training corpus that consists of the following sentences:
The Stormlands are ruled by House Baratheon.
Above the Stormlands we have the Crownlands.
The capital of the Crownlands is King's Landing.
The Westerlands are ruled by House Lannister.
In the centre we have the Riverlands.
Consider the test sentence:
In the centre we have the Stormlands.

(a) Compute the total number of unigrams, bigrams and trigrams of the training corpus (excluding $< /s >< s >$ transitions), and the dimension of the vocabulary.

(b) Compute the probability of the test utterance using a bigram language model without any type of smoothing.

(c) Compute the probability of the test utterance using a bigram language model with add-one smoothing.

(d) Which is the highest value of n, such that the probability of the n-gram model is non-zero for the sentence:

The Westerlands are ruled by House Baratheon.

11. (1.8) The following extract was produced by a speaker independent large vocabulary continuous speech recognition system for a broadcast news piece:
Várias potências mundiais reuniram-se em Viena para discutir a situação na Liga e mostraram-se dispostas a fornecer armas a um governo de unidade nacional líbio, para que o país possa fazer frente ao já vistas. Nesta revolução que depôs Kadhafi em dois mil e o se que o país vive uma situação política caótica e assiste a uma presença cada vez mais forte jihadistas no seu território.

The corresponding manual transcription is below:
Várias potências mundiais reuniram-se em Viena para discutir a situação na Líbia e mostraram-se dispostas a fornecer armas ao governo de unidade nacional líbio, para que o país possa fazer frente aos jihadistas. Desde a revolução que depôs Kadhafi em dois mil e onze que o país vive uma situação política caótica e assiste a uma presença cada vez mais forte dos jihadistas no seu território.

Ignoring punctuation and capitalization, compute the corresponding values of H (”correct”), D (”deletions”), S (”substitutions”), I (”insertions”), N (”total”), %Corr, %Acc, %WER, and %OOV.

H	D	S	I	N	% Corr	% Acc	% WER	% OOV

12. (1.5) Consider an audio file containing 10 minutes of speech, in which a keyword (out of a set of 7) is uttered 20 times. An automatic KWS system produces 25 keyword candidate hypotheses, 15 of them correspond to actual keyword detections. Compute (just indicate the operations) the false alarm (or insertion) ratio, the false rejection (or miss) ratio and the precision of the KWS system for this particular audio file. Assume that

1 word can be uttered every second.

Precision	
False alarm	
False rejection	

PART III: Chapter 5 & Chapter 6 True-False questions

13. (2.1) Classify as True (T) or False (F) on the left side of each item.

    (a) In SVM approaches, polynomial kernels are more prone to overfitting than linear ones.

    (b) PCA is a feature dimensionality reduction technique that needs labeled data for training.

    (c) Hybrid models combining neural networks and HMMs typically use an ergodic configuration for the latter.

    (d) Paralinguistic classification tasks typically use voice quality features.

    (e) The earlier VAD techniques explored both energy and duration thresholds.

    (f) Bootstrap methods for automatic alignment are used for unsupervised training of acoustic models.

    (g) DNN approaches typically map frames to senone posteriors, but CTC approaches map frames to phonemes or characters in a many-to-one mapping.

Name:	
Number:	
Group number:	